Categorical encoding of color in the brain

Article (Published Version)

Categorical encoding of color in the brain

Chris M. Bird[a,1], Samuel C. Berens[a], Aidan J. Horner[b,c], and Anna Franklin[a,d,1]

[a]School of Psychology and [d]The Sussex Colour Group, University of Sussex, Falmer BN1 9QH, United Kingdom; [b]University College London Institute of Cognitive Neuroscience, London WC1N 3AR, United Kingdom; and [c]University College London Institute of Neurology, London, WC1N 3BG, United Kingdom

Edited* by Paul Kay, Stanford University, Palo Alto, CA, and approved February 11, 2014 (received for review August 12, 2013)

The areas of the brain that encode color categorically have not yet been reliably identified. Here, we used functional MRI adaptation to identify neuronal populations that represent color categories irrespective of metric differences in color. Two colors were successively presented within a block of trials. The two colors were either from the same or different categories (e.g., "blue 1 and blue 2" or "blue 1 and green 1"), and the size of the hue difference was varied. Participants performed a target detection task unrelated to the difference in color. In the middle frontal gyrus of both hemispheres and to a lesser extent, the cerebellum, blood-oxygen level-dependent response was greater for colors from different categories relative to colors from the same category. Importantly, activation in these regions was not modulated by the size of the hue difference, suggesting that neurons in these regions represent color categorically, regardless of metric color difference. Representational similarity analyses, which investigated the similarity of the pattern of activity across local groups of voxels, identified other regions of the brain (including the visual cortex), which responded to metric but not categorical color differences. Therefore, categorical and metric hue differences appear to be coded in qualitatively different ways and in different brain regions. These findings have implications for the long-standing debate on the origin and nature of color categories, and also further our understanding of how color is processed by the brain.

categorization | functional magnetic resonance imaging | chromatic

Although color is continuous, humans can group the millions of discriminable colors into discrete categories, such as red, green, blue, and yellow (1). The origin and nature of such color categories has been extensively debated across the cognitive sciences (1–12). Traditionally, debate has focused on whether color categories are biologically constrained (2, 3), or whether they are arbitrary linguistic constructs that arise out of culture and communication (4). Alternative proposals include suggestions that color categories are a cognitive response to inequalities in perceptual color space (5), or that color categories are a property of the reflective surfaces of the visual environment (6). Also debated is the extent to which color categories affect how color is perceived. Some have argued that color categories affect the cognitive or attentional strategies of perceptual color judgments (7, 8), or that color categories affect early stages of color processing even when colors are not attended (9–12). However, others have argued that noncategorical sensory models of color encoding are sufficient to account for how color is perceived (13). The current investigation aims to contribute to the long-standing multidisciplinary debate on the origin and nature of color categories by identifying how color categories are represented in the brain.

Although there is some understanding of the areas of the brain involved in color vision, there is lack of clarity on where color is encoded categorically (14). It has been proposed (15)—but also refuted (16)—that clusters of color-preferring cells ("globs") in macaque posterior inferior temporal (IT) cortex represent the four "unique hues" (red, green, yellow, and blue) that all colors can be described in terms of. Neurons have also been identified in macaque IT that are more strongly excited during a color categorization task (is the color reddish or greenish?) compared with a color discrimination task (select the color that is the same as a sample color) (17). However, although IT neurons were more active when differentiating between red and green, their activity also discriminated within those categories; therefore, it was acknowledged that the neurons were not encoding color in a categorical manner. Perhaps related to this finding, optical imaging of macaque primary visual cortex has revealed a two-way spatial clustering of neural responses according to whether the L-M cone-contrast of a color is positive (e.g., reddish) or negative (e.g., greenish) (18). Furthermore, it was found that this distinction in neural activity correlated with the two-way classification of colors as "warm" or "cool," which appears to serve as a fault line in structure of the world's color lexicons (19). These important findings suggest a relationship between encoding of color at the visual cortex and categorization of color in language. However, the study does not suggest that encoding of color at the visual cortex is related to finer levels of color categorization that are subsumed within the relatively broad warm-cool categorical distinction (e.g., blue vs. green).

In humans, functional MRI (fMRI) studies have shown that the left posterior temporoparietal regions involved in color naming are activated when explicit identity judgments about color are made (e.g., are colors the same or different?) (20, 21), and that there is stronger activation in language networks when participants search for a colored target among different- rather than same-category colored distractors (22). In fact, the latter study also found "category effects" of greater activation for different- rather than same-category color search in 28 regions, including prefrontal regions and areas of the visual cortex (V2/V3). However, caution is required in interpreting these effects as categorical. Although such effects may appear to be related to color categories, inequalities in the color metric used to equate same- and different-category colors could well account for the effects (13, 23; this also applies to ref. 24). In addition, because search was faster for

Significance

Humans group the millions of discriminable colors into discrete categories, such as "blue" and "green." There has been much debate about where color categories come from; for example, whether color categories are inbuilt into the visual system. We use functional MRI to identify regions of the brain that categorize color. Color categories are encoded by regions of the frontal lobes, which also categorize other information (e.g., sounds). Interestingly, the visual cortex responds only to the size of color differences, but not color categories. We conclude that color categories occur at the level of attention rather than being inbuilt into the visual system. The findings shed light on how the brain categorizes information and how it processes color.

Author contributions: C.M.B., S.C.B., and A.F. designed research; C.M.B., S.C.B., and A.F. performed research; C.M.B., S.C.B., and A.J.H. analyzed data; and C.M.B. and A.F. wrote the paper.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

[1]To whom correspondence may be addressed. E-mail: chris.bird@sussex.ac.uk or anna.franklin@sussex.ac.uk.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1315275111/-/DCSupplemental.

different- rather than same-category target-distractor color pairs, categorical differences in blood-oxygen level-dependent (BOLD) activation are confounded by task difficulty: it may simply be the case that the identified regions are modulated by the difficulty of search rather than the categorical relationship of colored targets and distractors.

Three lines of evidence have suggested that color categories may influence processing at the level of early cortical visual regions. Electrophysiological studies have claimed that color categories modulate components of visually evoked event-related potentials that potentially arise from the visual cortex (10–12). However, it has been argued that inequalities in color metric or lack of clarity over the identity and origin of the event-related potential component complicate the interpretation of these findings (8). Using a very different approach, a behavioral-anatomical study found that the learning of novel color names was correlated with gray matter increases in V2/V3 (25). Although these results are encouraging, such increases in gray matter could be a result of general exposure and discrimination of the colors during the training period, rather than the specific categorical effect of color name learning. An fMRI study using an encoding model approach to describe the representation of color on the basis of multivariate patterns of brain responses showed that the hue tuning of areas VO1 and V4 undergoes categorical clustering, but only during color naming (26). This study suggests that the representation of color in these regions is flexible and responsive to top-down modulation contingent upon task demands. Nevertheless, it does not suggest that a categorical code for color is present at the visual cortex that is independent of explicit color naming, and the source of the top-down modulation remains unclear.

To summarize, although there have been attempts to identify the regions of the brain broadly related to categorical processing of color, no prior study has unequivocally identified which regions of the cortex encode color categorically in the absence of explicit color naming. The present investigation aims to reveal categorical encoding of color in the brain using the fMRI adaptation method. This method has been previously used to identify neuronal populations that encode other types of categories (e.g., ref. 27). The basic assumption of the method is that "neural adaptation reduces the BOLD response when successive stimuli activate the same subpopulation of neurons within a voxel, but not if they activate different subpopulations of neurons" (28). If a subpopulation of neurons encodes stimuli categorically, there should be a greater reduction in BOLD when successive stimuli are from the same category than when they are from different categories. Even if neural adaptation is not the precise mechanism of the reduction in BOLD response (see ref. 29 for a review of the different models of adaptation), the method remains a useful tool to study categorical representations.

We used a blocked design where participants viewed blocks of color stimulation during which two colors were presented in a successive fashion over 12 trials (Fig. 1 *A* and *B*). Participants were required to attend to colored stimuli, yet were not required to make explicit judgments about the identity or similarity of colors; therefore, category effects could not be confounded by the difficulty of any judgments made. To keep participants focused on viewing the colors, they were required to press a key when a target was present (a lighter patch in the center of one stimulus on a small percentage of the blocks). There were four colored stimuli, which were expected to be named in a subsequent naming test as one green and three blues. Colors were paired within a block, and across blocks each color was paired with every other color, including itself. On the basis of the intended color names (which were assessed by asking participants to name colors after the main experimental task), this gave six conditions where stimuli in a block were either same- or different-category, and where the difference in hue (hue is a perceptual dimension corresponding to dominant wavelength) between stimuli in a block was either absent (identical condition) or "small," "medium," or "large" (Fig. 1*D*). Therefore, the design included both a manipulation of the categorical relationship of colors in a block, and a metric manipulation of the size of the hue difference between colors in a block.

fMRI analyses aimed to identify regions of the brain where BOLD activation was greater for different- rather than same-category color blocks, and regions that had greater BOLD activity the larger the hue difference. Importantly, if neurons encode color categorically, then those neurons should respond to a categorical change in color (e.g., blue to green), but not to the size of the hue difference (e.g., small vs. medium color

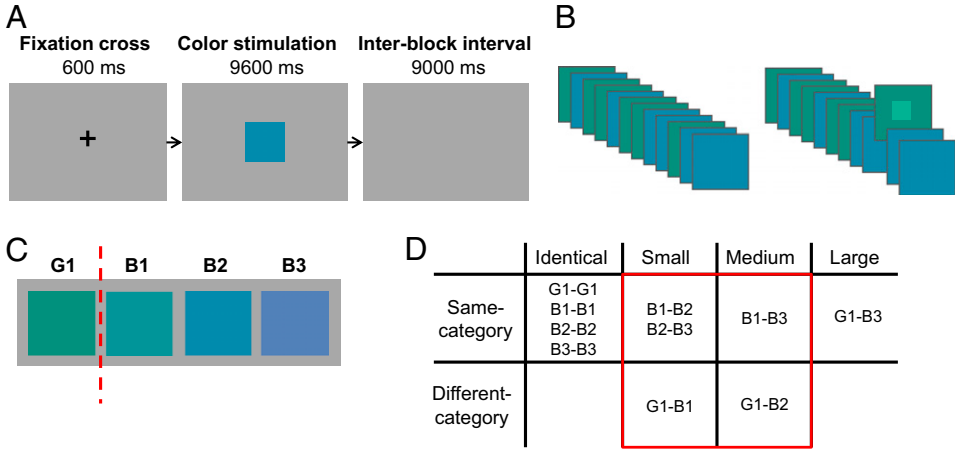

Fig. 1. fMRI adaptation stimuli and design. (*A*) Flow and time course of blocked design. During color stimulation a colored square was presented centrally on gray background (400 ms), 12 times separated by a gray background for 400 ms. (*B*) Two colors were presented six times each within a color stimulation block. On 12.5% blocks, one of the colored stimuli had a target that was a lighter square presented in the center of the colored stimulus (see right side of *B*). Participants were required to press a key when the target was detected, and blocks with targets were excluded from the analysis. (*C*) Print-rendered versions of the colors used. The dashed line indicates the blue-green lexical distinction made by the majority of the participants ($n = 17$): one stimulus was named green (G1) and three were named blue (B1/B2/B3). (*D*) Table of conditions. Across runs, all colors were paired with every other color, including itself, giving blocks where colors were either the same- or different-category, and where the size of the hue difference was absent (identical), small, medium, or large. The red box indicates the 2×2 design used in one analysis which aimed to identify regions of the brain that respond to categorical but not metric changes in color.

difference). Note that differences in discriminability of same- and different-category colors because of potential inequalities in color metric cannot account for a category effect in a region of the brain if neurons in that region do not respond to larger differences in discriminability resulting from explicit manipulation of the size of the hue difference. Alongside the adaptation analyses, which look at differences in activation within single voxels, we also performed analyses looking at the pattern of activation within a group of voxels [representational similarity analysis, RSA (30)]. The combination of adaptation and RSA approaches provided greater leverage to understand the neural basis of color categorization.

Broadly speaking, if color categories are represented at a fundamental sensory or perceptual level, then we predict categorical encoding of color in regions of the visual cortex. If color categories are only represented in language, we predict that only left posterior temporoparietal regions encode color categorically. If color categories arise as a result of top-down attentional processes, we would predict the frontal brain regions to be implicated. Of course, it is possible that color categories are represented at multiple stages of processing and that there are interactions between these brain regions.

Results

Preliminary Analyses. Analysis of the accuracy of target detection during the main experimental task confirmed that participants were attending to the stimuli throughout the experiment (7.8 of 8 targets detected on average; the poorest performing participant detected 5 of 8). Analysis of color naming verified that the majority of the participants (17 of 21) named the color stimuli as intended (as comprising one green and three blue stimuli) (Fig. 1*C*). The remaining four participants named the color stimuli on average as comprising two green and two blue colors, and data from these participants were not included in the following analyses of the other 17 participants, but were analyzed as a separate group in additional analyses (see *Additional Analysis of Four Extra Participants*, below).

Imaging Results. In all imaging analyses, we used a threshold of $P < 0.001$ uncorrected for multiple comparisons and had an extent threshold of 20 or more contiguous voxels. The first analysis compared all experimental blocks with an implicit baseline [unmodeled time, including fixation cross, interblock-interval (IBI), and rest period]. This analysis revealed no suprathreshold voxels, possibly because of the low demands of the task, the fact that an isoluminant gray background was present throughout the experiment, and that the baseline comprised an entirely unconstrained "rest" during which participants were free to engage in task-independent thought.

To independently investigate brain regions responding to changes in color category and to size of hue difference, we performed a 2 × 2 ANOVA on the fMRI data from the experimental blocks containing color pairs of small and medium hue difference (first factor, color category; second factor, size of hue difference) (Fig. 1*D*). Blocks where colors were identical or where there was a large hue difference were not included in this analysis, as there are no different-category pairs in the former blocks, nor any same-category pairs in the latter blocks (Fig. 1*D*).

There was a main effect of color category [blocks (G1-B1) and (G1-B2) > (B1-B2), (B2-B3) and (B1-B3)] in three brain regions: the left and right middle frontal gyrus (MFG) and the left cerebellum (Fig. 2 and Table 1). The sizes of both the left and right MFG clusters of activations are significant at the level of whole-brain family-wise error correction. These regions showed an increase in BOLD response for the different-category blocks compared with the same-category blocks, consistent with an fMRI adaptation effect, which predicts habituation of the BOLD response for repeated presentation of stimuli from the same color category.

Critically, there was no main effect of size of hue difference in any of the regions identified in the color category analysis, even at greatly reduced thresholds ($P < 0.01$, uncorrected for multiple comparisons). In fact, no regions showed a main effect of size of hue difference and no regions showed a significant interaction between color category and size of hue difference. Because it is risky to conclude no difference from failure to find a significant difference when null-hypothesis significance testing, we also analyzed the main effect of size of hue difference on MFG activation using a Bayesian model selection approach (31). This approach uses simple transformations of the sums of squares from the ANOVAs to generate Baysian information criterion probabilities (pBIC) of the null and alternative hypotheses given the data. For the average activity across the left MFG, the Bayes factor in favor of the null hypothesis over the main effect of size of hue difference was 3.26: pBIC(H1|D) = 0.235 and pBIC (H0|D) = 0.765. For the average activity across the right MFG, the Bayes factor in favor of the null hypothesis over the main effect of size of hue difference was 3.98: pBIC(H1|D) = 0.201 and pBIC(H0|D) = 0.799. These analyses therefore suggest that there is no effect of size of hue difference in both the left and right MFG. In each case, the probability of the alternative hypothesis is much smaller than that required for even a weak effect [where pBIC(H1|D) = 0.5–0.75 would indicate a weak effect (32, 33)], and a Bayes factor in support of the null hypothesis over 3 indicates substantial support for the null (32).

It is possible that some regions of the brain differentiated between the experimental blocks in terms of the patterns of

Table 1. Main effect of color category

Brain region	Brodmann area	Cluster size	MNI coordinates			*z*-score
			x	*y*	*z*	
Left MFG	9 /46	61*	−36	36	42	4.1
Right MFG	9	48*	33	33	45	3.8
			33	42	39	3.7
Left cerebellum	None	33	−15	−81	−51	4.6

Activations within the whole brain (statistical threshold $P < 0.001$ uncorrected, cluster size = 20 voxels).
*Cluster significant at $P < 0.05$, family-wise error-corrected for multiple comparisons across the whole brain.

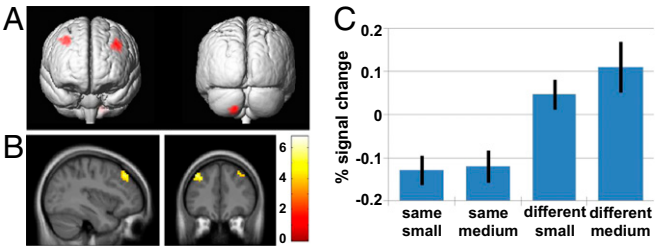

Fig. 2. Results of univariate analysis of the main effect of color category. (*A*) Main effect of color category in left and right MFG and cerebellum projected on to the group averaged structural scan ($P < 0.001$ uncorrected for multiple comparisons, extent threshold = 20 contiguous voxels). (*B*) The same contrast shown centered on the peak voxel in the left MFG. The colorbar represents the value of the *t*-statistic. (*C*) Mean percentage signal change for the four experimental block types compared with baseline shown for the peak voxel in the left MFG (error bars show SEMs). The baseline is an unconstrained "rest" period, and therefore it is the differences between experimental conditions that are critical rather than the absolute change relative to baseline.

activity across groups of voxels rather than in an overall change in activity. To investigate this theory, we performed three RSA (*Supporting Information*). The first RSA investigated whether any voxels exhibited a consistent pattern of activity during the experimental blocks versus the IBIs (Fig. S1*A*). The analysis revealed widespread regions of the visual cortex, which suggests these regions are representing the task stimuli despite not showing an overall increase in activity while performing the task (Fig. 3 *A* and *B*, Fig. S2 *A* and *B*, and Table S1).

The second analysis investigated whether the patterns of activity correlated more strongly between experimental blocks containing color pairs from the same category ("blue" blocks correlated with other "blue" blocks) versus from different categories ("blue" blocks correlated with "blue and green" blocks), while controlling for differences in the size of hue difference (Fig. S1*B*). No regions showed this effect. In the third RSA analysis we investigated whether the patterns of activity correlated more strongly between experimental blocks containing color pairs separated by a small hue difference versus a medium hue difference, while controlling for differences in color category (Fig. S1*C*). Three regions showed this effect (Fig. 3 *C* and *D*, Fig. S2 *C* and *D*, and Table S1). One of these regions, comprising 56 voxels, was in the visual cortex, and 14 of these voxels overlapped with the large visual cortical region that was identified in the first RSA analysis of the experimental trials versus IBIs. The regions identified in this third analysis are sensitive to the size of the hue difference, not in terms of an overall increase in activity, but in the pattern of activity across groups of voxels.

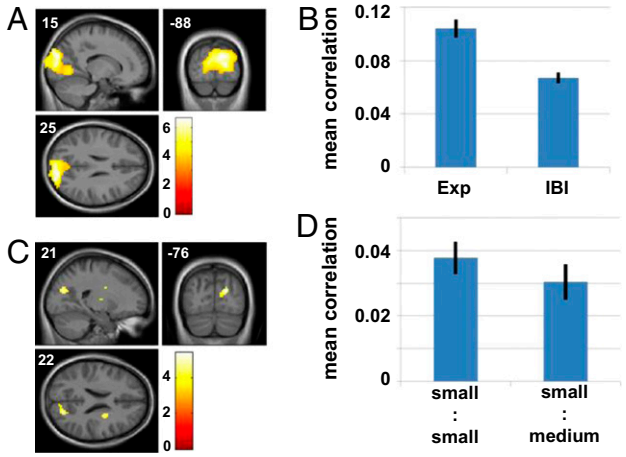

Fig. 3. RSA analyses of experimental blocks and color metric differences. Whole-brain "searchlight" analyses were carried out where the similarity in the pattern BOLD signal across 257 voxels in a moving sphere were compared against different phases of the experiment. The voxel at the center of the sphere was assigned the *t*-statistic for the comparison. (*A*) Brain regions where the similarity in patterns of BOLD signal was higher during experimental blocks than during IBIs, centered on the peak voxel showing the effect ($P < 0.001$ uncorrected). The colorbar represents the value of the *t*-statistic. (*B*) Bar graph showing the mean correlations between the experimental blocks and the interbloc intervals for the peak voxel identified in *A*, averaged across all participants (error bars show SEMs). Note that this is for illustrative purposes and does not represent an independent analysis. (*C*) Brain regions where the similarity in patterns of activations was greater during blocks with small rather than medium hue difference (color metric effect), centered on the peak voxel showing the effect ($P < 0.001$ uncorrected). The colored bar represents the value of the *t*-statistic. (*D*) Bar graph showing the mean correlations between blocks with small hue difference versus small and medium hue difference (error bars show SEMs). Note that this is for illustrative purposes and does not represent an independent analysis.

Additional Analysis of Four Extra Participants. It is possible that the MFG is responding to stimulus characteristics that might coincide with the location of the blue-green category boundary of our 17 participants. For example, one anonymous reviewer highlighted that the blue-green category boundary for the 17 participants coincides with whether the stimuli have an S-cone value above or below equal energy white (G1 is below and B1–B3 are above the S-cone value for equal energy white, and there is little variation in L-M cone values for the blue-green distinction). To account for our findings, the MFG would need to respond in a categorical manner to S-cone response, being modulated only by whether colors were above or below equal energy white and not being modulated by the size of the S-cone difference. We are not aware of any prior evidence that the S-cone signal can be extracted and treated in a categorical way based on the color's relationship to equal energy white.

We have some data that allow us to test this S-cone hypothesis, because 4 of the 21 participants named the colors differently to the rest of the sample. This subgroup of participants on average named the colors as two greens and two blues, and therefore perceived the blue-green category boundary to lie between B1 and B2. If the MFG is responding to the categorical status of colors rather than other stimulus characteristics (such as S-cone response), we would still predict a category effect within the MFG for this subgroup if the categorical status of color pairs is defined by their own color naming. Importantly, the different- and same-category blocks are different for the subgroup (where B1-B2 blocks are different-category and G1-B1 blocks are same-category) and the main group (where G1-B1 blocks are different-category and B1-B2 blocks are same-category). This color categorical interpretation of the effect was supported by mixed ANOVAs on the parameter estimates for the average response of all voxels in the left and right MFG clusters. Critically, there was no significant interaction between category (different vs. same) and group (main group and subgroup): left MFG: effect of category, $F(1, 19) = 12.1$, $P < 0.01$, effect of group, $F(1, 19) = 0.2$, $P = 0.69$, interaction, $F(1, 19) = 0.6$, $P = 0.43$; right MFG: effect of category, $F(1, 19) = 7.2$, $P < 0.02$, effect of group, $F(1, 19) = 0.2$, $P = 0.70$, interaction, $F(1, 19) = 0.8$, $P = 0.38$. Bayesian analysis (31) confirmed support for the null hypothesis over an interaction between category and group for the MFG on the left [Bayes factor in favor of null hypothesis = 3.19, $p(H0|D) = 0.761$, $p(H1|D) = 0.239$] and the right [Bayes Factor in favor of null hypothesis = 2.97, $p(H0|D) = 0.748$, $p(H1|D) = 0.252$]. This analysis therefore indicates that there is a category effect in the MFG when the categorical status of colors is defined according to each participant's color naming, irrespective of the location of the blue-green category boundary. In other words, the category effect at MFG is found for other patterns of color naming as well, and is therefore not restricted to certain stimuli.

In addition, if the effects in the main group analysis are because of perceived categorical membership and not other characteristics of the stimuli, then the subgroup of the four participants who name the colors differently to the main group will show a different pattern of BOLD response to blocks that are perceived as different- vs. same-category by the main group. We tested this by carrying out mixed ANOVAs on the parameter estimates for the average response of all voxels in the left and right MFG clusters, but this time defining blocks as different- or same-category according to the naming of the main group. Critically, in this analysis there was now a significant interaction between category (different- vs. same-category for the main group) and group (main group and subgroup) in both clusters: left MFG: effect of category, $F(1, 19) = 2.5$, $P = 0.13$, effect of group, $F(1, 19) = 0.1$, $P = 0.79$, interaction, $F(1, 19) = 34.2$, $P < 0.01$; right MFG: effect of category, $F(1, 19) = 6.3$, $P < 0.05$, effect of group, $F(1, 19) = 0.5$, $P = 0.48$, interaction, $F(1, 19) = 26.3$, $P < 0.01$. This analysis therefore indicates that the effect at MFG is for the blocks that

are perceived as different- vs. same-category by the main group, and is contingent on the perceived color categories of the colors in those blocks.

In sum, these post hoc analyses lend support to the notion that the regions of MFG identified in the main analysis are truly responding to perceived color categorical membership of the colors and the effect is not driven by other characteristics of the stimuli that may coincide with the location of the main group's blue-green color category boundary.

Discussion

We used fMRI to identify regions of the brain that independently code for differences in color category and the size of the hue difference. The MFG in both hemispheres showed stronger activation for different- vs. same-category color differences but was invariant to the size of the hue difference. No color categorical effects were observed in any visual cortical regions. However, there was a region of the visual cortex that was sensitive to the size of the hue difference, as revealed by the pattern of activity across voxels rather than overall changes in activity; the more similar the size of the hue difference between colors, the more similar the pattern of activity. An additional analysis investigated the similarity of patterns of firing for same- vs. different-category colors, while controlling for the size of the hue difference, but did not find any regions showing this effect. Therefore, categorical and metric hue differences appear to be coded in qualitatively different ways and in different brain regions. We discuss these findings in more detail below.

In the MFG and to a lesser extent, the cerebellum, BOLD activation was stronger for blue-green color differences than blue-blue, yet activation was not modulated by whether the hue difference was small or medium in size. The lack of a metric effect in these regions indicates that here, color is encoded in a purely categorical manner. Importantly, it also indicates that the category effect cannot be because of potential differences in discriminability of same- and different-category colors that might result from inequalities in the color metric. Even if such inequalities in color metric exist, they could not account for the category effect in the MFG, as larger differences in discriminability resulting from explicit manipulation of hue difference do not modulate activity in these regions. Therefore, although "category" effects of prior behavioral and neuroimaging studies might be a result of unequal same- and different-category hue difference rather than the categorical relationship between colors, the present study identifies an effect that is unequivocally categorical.

The most extensive regions to show category effects were the left and right MFG. We interpret the effect in this region as possibly reflecting a change detection process (explicit or implicit), operating at the level of conceptual categories. This interpretation may be underpinned by habituation of the firing of category-selective neurons during blocks, the common explanation for BOLD changes when using fMRI adaptation paradigms. Whatever their underlying neuronal origin, the category effects in the MFG arise even though participants are not required to make judgments about the identity of the colors, nor any other aspect of the hue difference. Note, no judgment of the hue difference between color pairs was required to detect the target, which was of a different lightness, and target blocks were not included in the analyses. Thus, categorical processing in the MFG appears to be automatic rather than effortful.

Additional support for the proposal that color category effects in the MFG reflect categorization at a conceptual level is that the region has been implicated in categorical processing in other domains, such as phonetic categorization (34), categorization of dot patterns (35), categorical spatial memory (36), semantic categories (37, 38), categorical uncertainty (39), and taxonomic categorization (40). Learned-object categories have also been found to be represented in a region homologous to the MFG in macaques (41). Taken together, the evidence supports a domain-general role for this region in categorization (see also ref. 34).

A categorical effect was also found in the cerebellum. The cerebellum has been implicated in a number of cognitive and affective processes, including previous color category studies (e.g., see figures 3*A* and 4*A* in ref. 21, and ref. 22). However, it is not clear whether these effects are simply because of its high density of connections to cortical regions where higher-order processing takes place, because the cerebellum is primarily associated with motor coordination (42, but see also ref. 43).

The findings of the present study have implications for the broad multidisciplinary debate on the origin and nature of color categories. Much of this debate has centered on whether color categories are biologically rooted in color perception (2) and the role of language in color categorical effects (4). The present study found no evidence of categorical encoding of color in classic visual and language regions of the brain. Although there is evidence that the very broad "warm-cool" category contrast is represented in the visual cortex (18), there has been no unequivocal evidence for encoding of finer categorical distinctions in the visual cortex that cannot be accounted for by top-down modulation resulting from explicit color naming (26). Furthermore, although language networks in the left temporal lobe involved in color naming are activated by explicit identity judgments about color (e.g., are the colors same or different?) (20), we find no evidence that these regions are involved when explicit judgments of hue difference are not required.

RSA investigated the correlation, or similarity, in the pattern of activity across local groups of voxels during different periods of the experiment. We found extensive regions of the visual cortex that showed higher correlations between the patterns of activities across voxels during experimental blocks than during IBIs, despite the fact that the overall activity was not significantly different during these periods. In a novel application of the RSA method, we investigated regions where pattern similarity was greater for small hue differences compared with larger differences. This process identified two right hemisphere regions: an area of the visual cortex, as well as an area involving the putamen and white matter of the right hemisphere. Activations in white matter, which has little energy requirements, are difficult to interpret. However, the identification of a visual cortical region, which partially overlaps the task-related activation in the first RSA analysis, is likely to be important. Combined with the lack of color category effects in the visual cortex in both the conventional univariate and the RSA analyses, our data suggest that the visual cortex is specialized for detecting metric differences in color. The fact that a metric effect was only detected using the RSA analysis suggests that metric differences may be coded for by shifting the weights in the firing patterns of local populations of neurons (detectable in the pattern of activity across voxels) rather than by overall increases or decreases in firing (which would result in a change in the level of activation).

The main contribution of the present study is to identify the brain regions that encode color in a categorical and not a metric manner. We have shown that color is encoded categorically, even when hue differences are irrelevant to an ongoing task. The only cortical regions to show the categorical effect were in the dorsolateral prefrontal cortex in both hemispheres, suggesting that automatic categorical encoding of color occurs at a conceptual stage of processing. Color categories may therefore not originate from computations at perceptual stages of color vision, but may rather arise from domain-general cognitive categorization processes.

Methods

Participants. Twenty-one participants (11 female, mean age 22.38 y, SD age = 2.18) gave written consent and were paid for participating, as approved by the Brighton and Sussex Medical School Research Ethics and Governance

Committee and the European Research Council Executive Agency Ethics Review Board. All were right-handed with normal or corrected-to-normal vision and reported to be in good health with no history of neurological disease. All participants had normal color vision as assessed by the Ishihara color plates (44).

Stimuli. Four colored stimuli from the blue-green region of CIELUV color space varied in Commission on Illumination (CIE) hue, with the size of the hue angle difference equated between adjacent stimuli (26.37°); CIE chroma (93.06) and lightness (L* =100) were kept constant. Stimuli were projected onto a screen in the MRI scanner from a calibrated video projector and the chromaticity coordinates of the screen-rendered colors were verified with a Minolta CS-100 colorimeter measuring from outside of the MRI bore via a system of mirrors. The CIE (1931) x, y chromaticity coordinates for the stimuli were: G1, $x = 0.258$, $y = 0.400$; B1, $x = 0.229$, $y = 0.339$; B2, $x = 0.225$, $y = 0.287$; B3, $x = 0.241$, $y = 0.252$. All stimuli had a luminance of $Y = 117.26$ cd/m^2, including the background gray ($x = 0.33$, $y = 0.33$).

Details of Procedure and Design. Participants performed 64 blocks; 58 were experimental blocks displaying six different types of color pairings: identical (for example, G1-G1), same-category small (B1-B2 or B2-B3), same-category medium (B1-B3), different-category small (G1-B1), different-category medium (G1-B2), and different-category large (G1-B3) (Fig. 1). There were also eight target-present blocks that occurred in all stimulus pairings. All blocks progressed in a pseudorandom order over the course of two runs separated by a short rest interval during which functional images were continuously acquired.

For each block, a black central fixation cross (1.3 cm^2) was presented for 0.6 s, followed by a 9.6-s period of color stimulation and then an IBI of 9 s. During nontarget blocks, 12 color squares (5 cm^2) were presented on a gray background centrally for 0.4 s each separated by the gray background alone for 0.4 s. The 12 color squares were six pseudorandomized presentations of two of the color stimuli (for example, G1 and B2), or 12 presentations of one of the color stimuli during identical color blocks. This design was similar for target blocks (12.5% of the blocks): two stimuli were alternated six times each except that, on one of the presentations, a lightly colored area inside a stimulus was visible ($Y = 90$ cd/m^2). Participants were instructed to respond to these targets with a button press.

After the main task, participants carried out a naming task to confirm the intended category membership of stimuli. Each stimulus was displayed in isolation on the same gray background and of the same size as the experimental task, and participants were asked to name the color as either "Blue" or "Green." This naming procedure was randomized and repeated three times for each stimulus.

Acquisition and Analysis of fMRI Time Series. Functional images were acquired on a Siemens Avanto 1.5 Tesla MRI scanner and analyzed by using SPM8, including standard preprocessing procedures (*Supporting Information*). fMRI time series were modeled by a general linear model including separate boxcar regressors for the experimental blocks (all pairings of color squares, eight in total). Target blocks and feedback periods were also modeled as separate boxcar regressors of "no interest." All regressors were convolved with the SPM hemodynamic response function. Data were high-pass filtered (cutoff period 128 s). Details of the first level (individual participant) univariate, and RSA analyses are given in the *Supporting Information*. Linear contrasts of coefficients for each participant were entered into second level random-effects analyses and considered statistically significant if they exceeded a threshold of $P < 0.001$ uncorrected for multiple comparisons and had an extent threshold of 20 or more contiguous voxels. Coordinates of brain regions are reported in Montreal Neurological Institute (MNI) space.

ACKNOWLEDGMENTS. We thank James Alvarez for assistance with rendering colors in the MRI scanner. This research was supported by a European Research Council Starting Grant (project "CATEGORIES") 283605 (to A.F.).

1. Regier T, Kay P (2009) Language, thought, and color: Whorf was half right. *Trends Cogn Sci* 13(10):439–446.
2. Regier T, Kay P, Cook RS (2005) Focal colors are universal after all. *Proc Natl Acad Sci USA* 102(23):8386–8391.
3. Franklin A, Davies IRL (2004) New evidence for infant color categories. *Br J Dev Psychol* 22:349–377.
4. Roberson D, Davies IRL, Davidoff J (2000) Color categories are not universal: Replications and new evidence from a stone-age culture. *J Exp Psychol Gen* 129(3):369–398.
5. Regier T, Kay P, Khetarpal N (2007) Color naming reflects optimal partitions of color space. *Proc Natl Acad Sci USA* 104(4):1436–1441.
6. Yendrikhozskij S (2001) Computing color categories from statistics of natural images. *J Imaging Sci Technol* 45(5):409–417.
7. Hanley JR, Roberson D (2011) Categorical perception effects reflect differences in typicality on within-category trials. *Psychon Bull Rev* 18(2):355–363.
8. Clifford A, et al. (2012) Neural correlates of acquired color category effects. *Brain Cogn* 80(1):126–143.
9. Holmes A, Franklin A, Clifford A, Davies IRL (2009) Neurophysiological evidence for categorical perception of color. *Brain Cogn* 69(2):426–434.
10. Thierry G, Athanasopoulos P, Wiggett A, Dering B, Kuipers JR (2009) Unconscious effects of language-specific terminology on preattentive color perception. *Proc Natl Acad Sci USA* 106(11):4567–4570.
11. Clifford A, Holmes A, Davies IRL, Franklin A (2010) Color categories affect pre-attentive color perception. *Biol Psychol* 85(2):275–282.
12. Mo L, Xu G, Kay P, Tan LH (2011) Electrophysiological evidence for the left-lateralized effect of language on preattentive categorical perception of color. *Proc Natl Acad Sci USA* 108(34):14026–14030.
13. Brown AM, Lindsey DT, Guckes KM (2011) Color names, color categories, and color-cued visual search: Sometimes, color perception is not categorical. *J Vis* 11(12):1–21.
14. Conway BR, et al. (2010) Advances in color science: From retina to behavior. *J Neurosci* 30(45):14955–14963.
15. Stoughton CM, Conway BR (2008) Neural basis for unique hues. *Curr Biol* 18(16): R698–R699.
16. Mollon JD (2009) A neural basis for unique hues? *Curr Biol* 19(11):R441–R442, author reply R442–R443.
17. Koida K, Komatsu H (2007) Effects of task demands on the responses of color-selective neurons in the inferior temporal cortex. *Nat Neurosci* 10(1):108–116.
18. Xiao Y, Kavanau C, Bertin L, Kaplan E (2011) The biological basis of a universal constraint on color naming: Cone contrasts and the two-way categorization of colors. *PLoS ONE* 6(9):e24994.
19. Lindsey DT, Brown AM (2006) Universality of color names. *Proc Natl Acad Sci USA* 103(44):16608–16613.
20. Tan LH, et al. (2008) Language affects patterns of brain activation associated with perceptual decision. *Proc Natl Acad Sci USA* 105(10):4004–4009.
21. Ikeda T, Osaka N (2007) How are colors memorized in working memory? A functional magnetic resonance imaging study. *Neuroreport* 18(2):111–114.
22. Ting Siok W, et al. (2009) Language regions of brain are operative in color perception. *Proc Natl Acad Sci USA* 106(20):8140–8145.
23. Witzel C, Gegenfurtner KR (2011) Is there a lateralized category effect for color? *J Vis* 11(12):16.
24. Walsh V, Kulikowski JJ, Butler SR, Carden D (1992) The effects of lesions of area V4 on the visual abilities of macaques: Colour categorization. *Behav Brain Res* 52(1):81–89.
25. Kwok V, et al. (2011) Learning new color names produces rapid increase in gray matter in the intact adult human cortex. *Proc Natl Acad Sci USA* 108(16):6686–6688.
26. Brouwer GJ, Heeger DJ (2013) Categorical clustering of the neural representation of color. *J Neurosci* 33(39):15454–15465.
27. van der Linden M, van Turennout M, Indefrey P (2010) Formation of category representations in superior temporal sulcus. *J Cogn Neurosci* 22(6):1270–1282.
28. Krekelberg B, Boynton GM, van Wezel RJ (2006) Adaptation: From single cells to BOLD signals. *Trends Neurosci* 29(5):250–256.
29. Grill-Spector K, Henson R, Martin A (2006) Repetition and the brain: Neural models of stimulus-specific effects. *Trends Cogn Sci* 10(1):14–23.
30. Kriegeskorte N, Mur M, Bandettini P (2008) Representational similarity analysis—Connecting the branches of systems neuroscience. *Front Syst Neurosci* 2:4.
31. Masson ME (2011) A tutorial on a practical Bayesian alternative to null-hypothesis significance testing. *Behav Res Methods* 43(3):679–690.
32. Raftery AE (1999) Bayes factors and BIC. *Sociol Methods Res* 27(3):411–427.
33. Dienes Z (2011) Bayesian versus orthodox statistics: Which side are you on? *Perspect Psychol Sci* 6(3):274–290.
34. Myers EB, Swan K (2012) Effects of category learning on neural sensitivity to non-native phonetic categories. *J Cogn Neurosci* 24(8):1695–1708.
35. Vogels R, Sary G, Dupont P, Orban GA (2002) Human brain regions involved in visual categorization. *Neuroimage* 16(2):401–414.
36. Slotnick SD, Moo LR (2006) Prefrontal cortex hemispheric specialization for categorical and coordinate visual spatial memory. *Neuropsychologia* 44(9):1560–1568.
37. Devlin JT, et al. (2002) Is there an anatomical basis for category-specificity? Semantic memory studies in PET and fMRI. *Neuropsychologia* 40(1):54–75.
38. Chan AH, et al. (2004) Neural systems for word meaning modulated by semantic ambiguity. *Neuroimage* 22(3):1128–1133.
39. Hansen KA, Hillenbrand SF, Ungerleider LG (2012) Effects of prior knowledge on decisions made under perceptual versus categorical uncertainty. *Frontiers in Neurosci* 6:1–10.
40. Sachs O, Weis S, Krings T, Huber W, Kircher T (2008) Categorical and thematic knowledge representation in the brain: Neural correlates of taxonomic and thematic conceptual relations. *Neuropsychologia* 46(2):409–418.
41. Freedman DJ, Riesenhuber M, Poggio T, Miller EK (2001) Categorical representation of visual stimuli in the primate prefrontal cortex. *Science* 291(5502):312–316.
42. Holmes G (1939) The cerebellum of man. *Brain* 30(1):466–488.
43. Strick PL, Dum RP, Fiez JA (2009) Cerebellum and nonmotor function. *Annu Rev Neurosci* 32:413–434.
44. Ishihara S (2004) *Ishihara's Tests for Colour Deficiency* (Kanehara Trading, Tokyo).

Supporting Information

Bird et al. 10.1073/pnas.1315275111

Imaging Protocol

Whole-brain T1-weighted structural scans were acquired with a 1–mm^3 resolution using a magnetization-prepared rapid gradient echo (MP-RAGE) pulse sequence. Blood-oxygen level-dependent (BOLD) sensitive, gradient-echo T2*-weighted scans were acquired using echo-planar imaging (EPI) comprising 36 contiguous axial slices (30° to AC-PC line; interleaved). The following imaging parameters were used: repetition time (TR) = 3300 ms, echo time (TE) = 50 ms, flip angle (FA) = 90°, slice thickness = 3 mm, in-plane resolution = 3 × 3 mm, acquisition matrix = 64 × 64, and field of view (FoV) = 192 × 192 mm. To allow for T1 equilibrium, the first four EPI volumes were acquired before the task started and then discarded. In addition, a field map using a double-echo FLASH sequence was recorded for distortion correction of the acquired EPI images (see below).

Image Preprocessing

All preprocessing and statistical analyses were performed using SPM8 (www.fil.ion.ucl.ac.uk/spm). EPI scans were spatially realigned to the first scan in the times series and were corrected for distortions based on the field map (1) and the interaction of motion and distortion using the Unwarp routines in SPM (1, 2). Each participant's structural scan was coregistered to a mean image of their realigned EPI scans and then used to calculate transformation parameters for normalizing the EPI scans to the Montreal Neurological Institute (MNI) template brain. Finally, the normalized EPI images were spatially smoothed with an isotropic 8-mm full-width at half-maximum (FWHM) Gaussian kernel. Structural scans were normalized and averaged using the DARTEL toolbox (3), and these mean images were used to display the data.

Univariate Analyses

The functional MRI (fMRI) time series were modeled by a general linear model detailed in the main text (*Results*).

The first analysis compared the pooled data from all of the experimental blocks with an implicit baseline [unmodeled time, including fixation cross, interblock-interval (IBI) and rest period].

The second analysis was a 2 × 2 ANOVA with the factors color category and size of hue difference carried out on the parameter estimates for the regressors of the appropriate experimental blocks (*Results*).

Linear contrasts of coefficients for each participant were entered into second level random-effects analyses and considered statistically significant if they exceeded a threshold of $P < 0.001$ uncorrected for multiple comparisons, and had an extent threshold of 20 or more contiguous voxels.

Representational Similarity Analyses

Unsmoothed realigned data were used in the representational similarity analysis (RSA) analyses for individual participants.

In the first analysis we investigated which brain regions showed higher correlated patterns of activity during experimental blocks than during IBIs. Specifically, we tested the prediction illustrated in Fig. S1*A*.

In a first-level general linear model for each individual we modeled all experimental blocks (56 in total) and half of the IBIs (28 in total; the remainder contributed to the unmodeled time to allow an accurate estimate of the mean activity across the session), as well as a single regressors for all target containing blocks and a single regressor for all feedback periods. This process resulted in the estimation of 86 separate regressors. Contrast images for each regressor (compared with unmodeled time) were then calculated to obtain an image of the "pattern" of activity for all of the modeled experimental periods.

A whole-brain searchlight procedure was used to correlate the pattern of activity across local groups of voxels between different experimental periods. The radius of the searchlight volume was 12 mm, resulting in a searchlight volume of 257 voxels (which was slightly elliptical in the *z* axis, because of a 0.75-mm gap between slices in the *z*-plane). For each searchlight, a 56 × 56 matrix was calculated, correlating the pattern of activity (across the voxels within the searchlight) between each of the 56 patterns (all 28 modeled IBIs and 28 of the modeled experimental blocks). This correlation matrix was then assigned to the voxel at the center of the searchlight. In a second step, we compared these correlation matrices with predefined models. These models were 56 × 56 matrices of the correlations expected between conditions (4). In this analysis we predicted more similarity between patterns of activity during any of the color blocks relative to when no color is presented, resulting in the model shown in Fig. S2*A*. All conditions relating to color blocks (28 × 28 events) were assigned a value of 1, whereas all conditions relating to IBI periods (no color stimulation; 28 × 28 events) were assigned a value of −1. We correlated this model with the actual correlation matrix from the RSA analysis (Fig. S2*B*) for each voxel in the brain to assess whether the pattern of correlations across conditions was captured by our specific model. The resultant β-value estimated by correlating the data with the model was assigned to that voxel. This process resulted in a whole-brain image of β-values for each participant.

These images were normalized to MNI space and smoothed with an isotropic 6-mm FWHM Gaussian kernel and were then analyzed in a second level group analysis using standard procedures implemented in SPM8. A one-sample *t* test was used to identify voxels that showed a significant effect (i.e., where the pattern of activity across conditions was captured by our model). We used a threshold of $P < 0.001$ (uncorrected for multiple comparisons) and an extent threshold of 20 contiguous voxels. The results of this analysis are shown in Fig. 3 and Table S1.

For the RSA analysis of color category, each individuals' fMRI data were preprocessed in the same way as described above, but the first-level model for each participant included all experimental blocks (56 in total), all target blocks (8 in total) and all feedback blocks (8 in total) but did not model the IBIs. A searchlight volume with radius 12 mm was again used. To test the prediction shown in Fig. S1*B*, we investigated the similarity in patterns of activity between experimental blocks that contained color pairs that were from the same color category, with patterns of activity between experimental blocks that contained color pairs that were from different color categories. Specifically, we investigated whether the correlation between the B1-B2 color pair blocks with B1-B2 and B2-B3 color pair blocks was larger than the correlation between the B1-B2 color pair blocks with G1-B1 and G1-B2 color pair blocks. The data were analyzed using the same steps and using the same statistical threshold as described above.

For the RSA analyses of size of hue difference, a similar analysis was carried out to the category analysis detailed above. To test the prediction shown in Fig. S1*C*, we investigated the similarity in patterns of activity between experimental blocks that contained color pairs that had a small hue difference with patterns of activity between experimental blocks that contained color pairs of a larger hue difference. Specifically, we investigated

whether the correlation between the B1-B2 color pair blocks with G1-B1 and B1-B2 color pair blocks was larger than the correlation between the B1-B2 color pair blocks with G1-B2 and B1-B3 color pair blocks. The data were analyzed using the same steps and using the same statistical threshold as described above.

1. Hutton C, et al. (2002) Image distortion correction in fMRI: A quantitative evaluation. *Neuroimage* 16(1):217–240.
2. Andersson JL, Hutton C, Ashburner J, Turner R, Friston K (2001) Modeling geometric deformations in EPI time series. *Neuroimage* 13(5):903–919.
3. Ashburner J (2007) A fast diffeomorphic image registration algorithm. *Neuroimage* 38(1):95–113.
4. Kriegeskorte N, Mur M, Bandettini P (2008) Representational similarity analysis—Connecting the branches of systems neuroscience. *Front Syst Neurosci* 2:4.

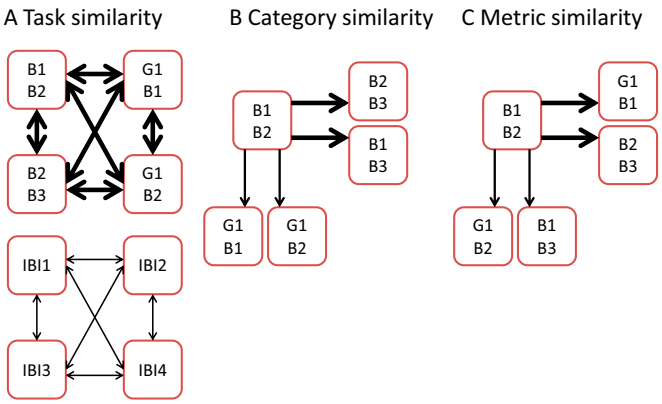

Fig. S1. Models tested for the representational similarity analyses. (*A*) Main effect of experimental task. Brain regions involved in the task will show greater correlations between patterns of activity during experimental trials than during IBIs. Note, only a subset of experimental blocks and IBIs are shown. (*B*) Color categorical effect. Experimental blocks containing B1-B2 color pairs (same-category) will show greater correlated patterns of activity with blocks containing B2-B3 and B1-B3 color pairs (same-category) than with blocks containing G1-B1 and G1-B2 color pairs (different-category). (*C*) Metric hue difference effect. Experimental blocks containing B1-B2 color pairs (small hue difference) will show greater correlated patterns of activity with blocks containing G1-B1 and B2-B3 color pairs (small hue difference) than with blocks containing G1-B2 and B1-B3 color pairs (medium hue difference).

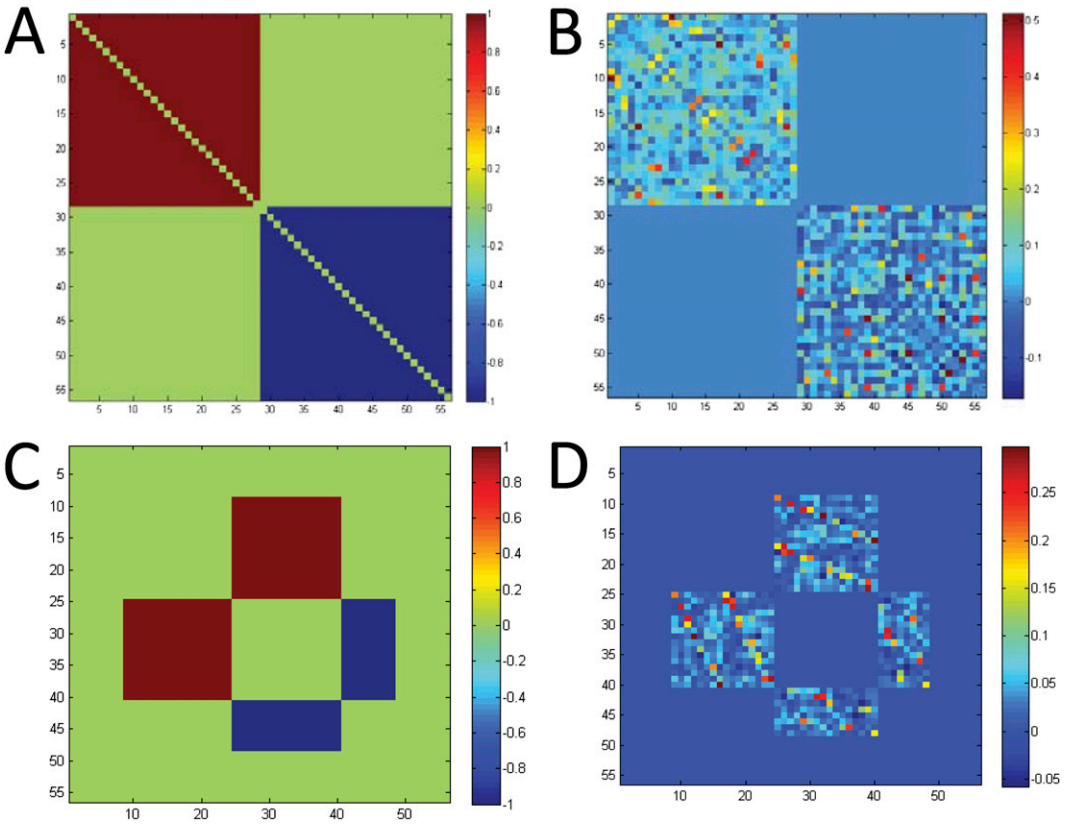

Fig. S2. (*A*) RSA model for the main effect of experimental task. (*Upper Left*) Experimental trials where a high correlation (red) in the pattern of activity in the visual cortex is expected; (*Lower Right*) IBIs where a lower correlation (blue) in the pattern of activity in visual cortex is expected. Correlations between identical trials (main diagonal) and between experimental trails and IBIs are not modeled and set to zero. (*B*) Color map showing the strength of the correlations between the blocks modeled in *A* for the peak voxel identified in this analysis (see main text), averaged across all participants. (*C*) Model for the metric hue difference effect; correlations between different experimental blocks with a small hue difference are expected to have a high correlation, correlations between blocks with a small hue difference, and blocks with a medium hue difference are expected to have a low correlation. (*D*) Color map showing the strength of correlations between the blocks modeled in *C* for the peak voxel identified in this analysis, which was in visual cortex (see main text). "Stripes" of higher correlations are between blocks that were closer in time, possibly caused by slow movements during the scanning session.

Table S1. RSA analyses

Brain region	Brodmann area	Cluster size	MNI coordinates			*z*-score
			x	*y*	*z*	
Correlation within experimental blocks > within IBI						
Right/left calcarine cortex, cuneus, lingual gyrus, middle occipital gyrus, superior occipital gyrus	17/18	2,387*	15	−88	25	4.5
			18	−88	10	4.5
			21	−94	28	4.5
Right/left precuneus	7	65	3	−64	61	3.4
			0	−58	52	3.3
			−9	−64	49	3.1
Left middle occipital gyrus	19	21	−48	−79	7	3.2
Correlation between small hue difference blocks > large hue difference blocks						
Right calcarine cortex, cuneus, superior occipital gyrus	18/19	56	21	−76	22	4.0
Right putamen/white matter	48/none	147*	24	−7	7	4.0
			30	2	10	3.8
			27	−4	28	3.7

Regions where the contrast between correlations is significant across the whole brain (statistical threshold $P < 0.001$ uncorrected, cluster size = 20 voxels).
*Cluster significant at $P < 0.05$, family-wise error-corrected for multiple comparisons across the whole brain.